

Convergence analysis in convex regularization depending on the smoothness degree of the penalizer

Erdem Altuntac

Institute for Numerical and Applied Mathematics, University of Göttingen, Lotzestr. 16-18, D-37083, Göttingen, Germany

E-mail: e.altuntac@math.uni-goettingen.de

Abstract.

The problem of minimization of the least squares functional with a smooth, lower semi-continuous, convex penalizer $J(\cdot)$ is considered to be solved. Over some compact and convex subset Ω of the Hilbert space \mathcal{H} , the regularizer is implicitly defined as $J(\cdot) : \mathcal{C}^k(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$ where $k \in \{1, 2\}$. So the cost functional associated with some given linear, compact and injective forward operator $\mathcal{T} : \Omega \subset \mathcal{H} \rightarrow \mathcal{H}$,

$$F_\alpha(\cdot, f^\delta) := \frac{1}{2} \|\mathcal{T}(\cdot) - f^\delta\|_{\mathcal{H}}^2 + \alpha J(\cdot),$$

where f^δ is the given perturbed data with its perturbation amount δ in it. Convergence of the regularized optimum solution $\varphi_{\alpha(\delta)} \in \arg \min F_\alpha(\varphi, f^\delta)$ to the true solution φ^\dagger is analysed depending on the smoothness degree of the penalizer, *i.e.* the cases $k \in \{1, 2\}$ in $J(\cdot) : \mathcal{C}^k(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$. In both cases, we define such a regularization parameter that is in cooperation with the condition

$$\alpha(\delta, f^\delta) \in \{\alpha > 0 \mid \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \tau\delta\},$$

for some fixed $\tau \geq 1$. In the case of $k = 2$, we are able to evaluate the discrepancy $\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \tau\delta$ with the Hessian Lipschitz constant L_H of the functional $F_\alpha(\cdot, f^\delta)$.

Keywords. convex regularization, Bregman divergence, Hessian Lipschitz constant, discrepancy principle.

1. Introduction

In this work, over some compact and convex subset Ω of the Hilbert space \mathcal{H} , we consider solving formulate our main variational minimization problem,

$$\arg \min_{\Omega \subset \mathcal{H}} \left\{ F_\alpha(\cdot, f^\delta) := \frac{1}{2} \|\mathcal{T}(\cdot) - f^\delta\|_{\mathcal{H}}^2 + \alpha J(\cdot) \right\}. \quad (1.1)$$

Here, $J(\cdot) : \mathcal{C}^k(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$, for $k = \{1, 2\}$ is convex and $\alpha > 0$ is the regularization parameter. Following [10, 13, 18], we construct the parametrized solution $\varphi_{\alpha(\delta)}$ for the problem (1.1) satisfying

- (i) For any $f \in \mathcal{H}$ there exists a solution $\varphi_\alpha \in \mathcal{H}$ to the problem (1.1);
- (ii) For any $f \in \mathcal{H}$ there is no more than one $\varphi_\alpha \in \mathcal{H}$;
- (iii) Convergence of the regularized solution φ_α to the true solution φ^\dagger must depend on the given data, *i.e.*

$$\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|_{\mathcal{H}} \rightarrow 0 \text{ as } \alpha(\delta) \rightarrow 0 \text{ for } \delta \rightarrow 0$$

whilst

$$\|f^\dagger - f^\delta\| \leq \delta$$

where $f^\dagger \in \mathcal{H}$ is the true measurement and δ is the noise level.

What is stated by ‘(iii)’ is that when the given measurement f^δ lies in some δ –ball centered at the true measurement f^\dagger , $\mathcal{B}_\delta(f^\dagger)$, then the expected solution must lie in the corresponding $\alpha(\delta)$ ball. It is also required that this solution $\varphi_{\alpha(\delta)}$ must depend on the data f^δ . Therefore, we are always tasked with finding an approximation of the unbounded inverse operator $\mathcal{T}^{-1} : \mathcal{R}(T) \rightarrow \mathcal{H}$ by a bounded linear operator $R_\alpha : \mathcal{H} \rightarrow \mathcal{H}$.

DEFINITION 1.1 (Regularization operator). *[10, Definition 4.3], [20, Theorem 2.2] Let $\mathcal{T} : \mathcal{H} \rightarrow \mathcal{H}$ be some given linear injective operator. Then a family of bounded operators $R_\alpha : \mathcal{H} \rightarrow \mathcal{H}$, $\alpha > 0$, with the property of pointwise convergence*

$$\lim_{\alpha \rightarrow 0} R_\alpha \mathcal{T} \varphi^\dagger = \varphi^\dagger \tag{1.2}$$

is called a regularization scheme for the operator \mathcal{T} . The parameter α is called regularization parameter.

As alternative to well established Tikhonov regularization, [21, 22], studying convex variational regularization with any penalizer $J(\cdot)$ has become important over the last decade. Introducing a new image denoising method named as *total variation*, [24], is commencement of this study. Application and analysis of the method have been widely carried out in the communities of inverse problems and optimization, [1, 2, 4, 7, 8, 9, 11, 12, 25]. Particularly, formulating the minimization problem as variational problem and estimating convergence rates with variational source conditions has also become popular recently, [6, 15, 16, 17, 20]. Different from available literature, we take into account one fact; for some given measurement f^δ with the noise level δ and forward operator \mathcal{T} , the regularized solution $\varphi_{\alpha(\delta)}$ to the problem (1.1) should satisfy $\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \tau\delta$ for some fixed $\tau \geq 1$. With this fact, we manage to obtain tight convergence rates for $\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|$, and we can carry out this analysis for a general smooth, convex penalty $J(\cdot) \in \mathcal{C}^k(\Omega, \mathcal{H})$ for the cases $k = \{1, 2\}$. We will be able to quantify the tight convergence rates under the assumption that $J(\cdot)$ is defined over $\mathcal{C}^k(\Omega, \mathcal{H})$ space for $k \in \{1, 2\}$. To be more specific, we will observe that rule for the choice of regularization parameter $\alpha(\delta)$ must contain Lipschitz constant in addition to the noise level δ . That is, when $k = 2$, we will need \mathcal{C}^{2+} class.

2. Notations and prerequisite knowledge

Let $\mathcal{C}(\Omega)$ be the space of continuous functions on the compact domain Ω . Then, $\mathcal{C}^k(\Omega)$ function space

$$\mathcal{C}^k(\Omega) := \{\varphi \in \mathcal{C}(\Omega) : \nabla^{(k)}\varphi \in \mathcal{C}(\Omega)\}.$$

Addition to traditional \mathcal{C}^k spaces, we will need to address \mathcal{C}^{k+} for the purpose of convergence analysis. In general for an open set $O \subset \mathbb{R}^N$, a mapping $\mathcal{P} : O \rightarrow \mathbb{R}^N$ is said to be of *class* \mathcal{C}^{k+} if it is of class \mathcal{C}^k and k th partial derivatives are not just continuous but strictly continuous on O , [23, pp. 355]. Then, for a smooth and convex functional $J(\varphi)$ defined over $\mathcal{C}^k(\Omega, \mathcal{H})$, there exists Lipschitz constant \tilde{L} such that

$$\|\nabla^{(k)}J(\varphi) - \nabla^{(k)}J(\Psi)\| \leq \tilde{L}\|\varphi - \Psi\|. \quad (2.1)$$

When $k = 1$, by \tilde{L} we denote well-known Lipschitz constant L . When $k = 2$, \tilde{L} will be Hessian Lipschitz L_H , [14].

Over some compact and convex domain $\Omega \subset \mathcal{H}$, variational minimization problem is formulated as such,

$$\arg \min_{\varphi \in \mathcal{H}} \left\{ F_\alpha(\cdot, f^\delta) := \frac{1}{2}\|\mathcal{T}(\cdot) - f^\delta\|_{\mathcal{H}}^2 + \alpha J(\cdot) \right\} \quad (2.2)$$

with its penalty $J(\cdot) : \mathcal{C}^k(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$, where $k = \{1, 2\}$, and $\alpha > 0$ is the regularization parameter. Another dual minimization problem to (2.2) is given by

$$J(\cdot) \rightarrow \min_{\mathcal{H}}, \text{ subject to } \|\mathcal{T}(\cdot) - f^\delta\| \leq \delta. \quad (2.3)$$

In the Hilbert scales, it is known that the solution of the penalized minimization problem (2.2) equals to the solution of the constrained minimization problem (2.3), [6, Subsection 3.1]. The regularized solution $\varphi_{\alpha(\delta)}$ of the problem (2.2) satisfies the following first order optimality conditions,

$$\begin{aligned} 0 &= \nabla F_\alpha(\varphi_{\alpha(\delta)}) \\ 0 &= \mathcal{T}^*(\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta) + \alpha(\delta)\nabla J(\varphi_{\alpha(\delta)}) \\ \mathcal{T}^*(f^\delta - \mathcal{T}\varphi_{\alpha(\delta)}) &= \alpha(\delta)\nabla J(\varphi_{\alpha(\delta)}). \end{aligned} \quad (2.4)$$

In this work, the radii δ of the $\alpha(\delta)$ ball are estimated, by means of the Bregman divergence, with potential $J(\cdot) : \mathcal{C}^1(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$. The choice of regularization parameter $\alpha(\delta)$ in this work does not require any *a priori* knowledge about the true solution. We always work with perturbed data f^δ and introduce the rates according to the perturbation amount δ .

2.1. Bregman divergence

We will be able to quantify the rate of the convergence of $\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|$ by means of different formulations of the Bregman divergence. Following formulation emphasizes the functionality of the Bregman divergence in proving the norm convergence of the minimizer of the convex minimization problem to the true solution.

DEFINITION 2.1 (Total convexity and Bregman divergence). [5, Def.1]

Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ be a smooth and convex functional. Then Φ is called totally convex in $u^* \in \mathcal{H}$, if, for $\nabla\Phi(u^*)$ and $\{u\}$, it holds that

$$D_\Phi(u, u^*) = \Phi(u) - \Phi(u^*) - \langle \nabla\Phi(u^*), u - u^* \rangle \rightarrow 0 \Rightarrow \|u - u^*\|_{\mathcal{H}} \rightarrow 0$$

where $D_\Phi(u, u^*)$ represents the Bregman divergence.

It is said that Φ is q -convex in $u^* \in \mathcal{H}$ with a $q \in [2, \infty)$, if for all $M > 0$ there exists a $c^* > 0$ such that for all $\|u - u^*\|_{\mathcal{H}} \leq M$ we have

$$D_\Phi(u, u^*) = \Phi(u) - \Phi(u^*) - \langle \nabla\Phi(u^*), u - u^* \rangle \geq c^* \|u - u^*\|_{\mathcal{H}}^q. \quad (2.5)$$

Throughout our norm convergence estimations, we refer to this definition for the case of 2-convexity. We will also study different formulations of the Bregman divergence. We introduce these different formulations below.

REMARK 2.2 (Different formulations of the Bregman divergence). Let $\varphi_{\alpha(\delta)}, \varphi^\dagger$ defined on Ω respectively be the regularized and the true solutions of the problem (2.2). Then we give the following definitions of the Bregman divergence;

- Bregman distance associated with the cost functional $F(\cdot)$:

$$D_F(\varphi_{\alpha(\delta)}, \varphi^\dagger) = F(\varphi_{\alpha(\delta)}) - F(\varphi^\dagger) - \langle \nabla F(\varphi^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle, \quad (2.6)$$

- Bregman distance associated with the penalty $J(\cdot)$:

$$D_J(\varphi_{\alpha(\delta)}, \varphi^\dagger) = J(\varphi_{\alpha(\delta)}) - J(\varphi^\dagger) - \langle \nabla J(\varphi^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle \quad (2.7)$$

- Bregman distance associated with the misfit term $G_\delta(\cdot, f^\delta) := \frac{1}{2} \|\mathcal{T}(\cdot) - f^\delta\|^2$:

$$D_{G_\delta}(\varphi_{\alpha(\delta)}, \varphi^\dagger) = \frac{1}{2} \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|^2 - \frac{1}{2} \|\mathcal{T}\varphi^\dagger - f^\delta\|^2 - \langle \nabla G_\delta(\varphi^\dagger, f^\delta), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle \quad (2.8)$$

Reader may also refer to Appendix A for further properties of the Bregman divergence. In fact, another similar estimation to (2.5), for $q = 2$, can also be derived by making further assumption about the functional Φ one of which is strong convexity with modulus c , [3, Definition 10.5]. Below is this alternative way of obtaining (2.5) when $q = 2$.

PROPOSITION 2.3. Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ be $\Phi \in \mathcal{C}^2(\mathcal{H})$ is strongly convex with modulus of convexity $c > 0$, i.e. $\nabla^2\Phi \succ cI$, then

$$D_\Phi(u, v) > c\|u - v\|^2 + \mathcal{O}(\|u - v\|^2). \quad (2.9)$$

Proof. Let us begin with considering the Taylor expansion of Φ ,

$$\Phi(u) = \Phi(v) + \langle \nabla \Phi(v), u-v \rangle + \frac{1}{2} \langle \nabla^2 \Phi(v)(u-v), u-v \rangle + \mathcal{O}(\|u-v\|^2). \quad (2.10)$$

Then the Bregman divergence

$$\begin{aligned} D_\Phi(u, v) &= \Phi(u) - \Phi(v) - \langle \nabla \Phi(v), u-v \rangle \\ &= \langle \nabla \Phi(v), u-v \rangle + \frac{1}{2} \langle \nabla^2 \Phi(v)(u-v), u-v \rangle + \mathcal{O}(\|u-v\|^2) - \langle \nabla \Phi(v), u-v \rangle \\ &= \frac{1}{2} \langle \nabla^2 \Phi(v)(u-v), u-v \rangle + \mathcal{O}(\|u-v\|^2). \end{aligned}$$

Since $\Phi(\cdot)$ is strictly convex, due to strong convexity and $\Phi \in \mathcal{C}^2(\mathcal{H})$, hence one obtains that

$$D_\Phi(u, v) > c\|u-v\|^2 + \mathcal{O}(\|u-v\|^2), \quad (2.11)$$

where c is the modulus of convexity. □

Above, in (2.8), we have set $\Phi := G_\delta(\cdot, f^\delta)$. In this case, one must assume even more than stated about the existence of the modulus of convexity c . These assumptions can be formulated in the following way. Suppose that there exists some measurement f^δ lying in the δ -ball $\mathcal{B}_\delta(f^\dagger)$ for all $\delta > 0$ small enough such that the followings hold,

$$0 < c_\delta \leq c_{f^\delta}, \quad (2.12)$$

$$0 < \underline{c} \leq c_\delta, \text{ for all } \delta > 0. \quad (2.13)$$

Then $G_\delta(\cdot, f^\delta)$ is 2-convex and according to Proposition 2.3,

$$D_{G_\delta}(\varphi_{\alpha(\delta)}, \varphi^\dagger) > c_{f^\delta} \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2 + \mathcal{O}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2), \quad (2.14)$$

Addition to the traditional definition of Bregman divergence in (2.5), *symmetrical Bregman divergence* is also given below, [16, Definition 2.1],

$$D_\Phi^{\text{sym}}(u, u^*) := D_\Phi(u, u^*) + D_\Phi(u^*, u). \quad (2.15)$$

With symmetrical Bregman divergence having formulated, following from the Definition 2.1, we give the last proposition for this chapter.

PROPOSITION 2.4. [16, as appears in the proof of Theorem 4.4] *Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ be a smooth and q -convex functional. Then there exist positive constants $c^*, c > 0$ such that for all $\|u - u^*\|_{\mathcal{H}} \leq M$ we have*

$$\begin{aligned} D_\Phi^{\text{sym}}(u, u^*) &= \langle \nabla \Phi(u^*) - \nabla \Phi(\tilde{u}), u - u^* \rangle \\ &\geq (c^* + c) \|u - u^*\|_{\mathcal{H}}^2. \end{aligned} \quad (2.16)$$

Proof. Proof is a straightforward result of the estimation in (2.5) and the symmetrical Bregman divergence definition given by (2.15). \square

2.2. Appropriate regularization parameter with discrepancy principle

A regularization parameter α is admissible for δ when

$$\|\mathcal{T}\varphi_\alpha - f^\delta\| \leq \tau\delta \quad (2.17)$$

for some fixed $\tau \geq 1$. We seek a rule for choosing $\alpha(\delta)$ as a function of δ such that (2.17) is satisfied and

$$\alpha(\delta) \rightarrow 0, \text{ as } \delta \rightarrow 0.$$

Following [13, Eq. (4.57) and (4.58)], [19, Definition 2.3], in order to obtain tight rates of convergence of $\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|$ we define $\alpha(\delta, f^\delta)$ such that

$$\alpha(\delta, f^\delta) \in \{\alpha > 0 \mid \|\mathcal{T}\varphi_\alpha - f^\delta\| \leq \tau\delta, \text{ for all given } (\delta, f^\delta)\}. \quad (2.18)$$

The strong relation between the discrepancy $\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|$ and the norm convergence of $\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|$ can be formulated in the following lemma.

LEMMA 2.5. *Let $\mathcal{T} : \mathcal{H} \rightarrow \mathcal{H}$ be a linear and compact operator. Denote by $\varphi_{\alpha(\delta)}$ the regularized solution and by φ^\dagger the true solution to the problem (2.2). Then*

$$\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \delta + \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \|\mathcal{T}^*\|, \quad (2.19)$$

where the noisy data f^δ to the true data f^\dagger both satisfy $\|f^\delta - f^\dagger\| \leq \delta$ for sufficiently small amount of noise δ .

Proof. Desired result follows from the following straightforward calculations,

$$\begin{aligned} \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|^2 &= \langle \mathcal{T}\varphi_{\alpha(\delta)} - f^\delta, \mathcal{T}\varphi_{\alpha(\delta)} - f^\delta \rangle \\ &= \langle \mathcal{T}\varphi_{\alpha(\delta)} - f^\dagger + f^\dagger - f^\delta, \mathcal{T}\varphi_{\alpha(\delta)} - f^\delta \rangle \\ &= \langle \mathcal{T}\varphi_{\alpha(\delta)} - f^\dagger, \mathcal{T}\varphi_{\alpha(\delta)} - f^\delta \rangle + \langle f^\dagger - f^\delta, \mathcal{T}\varphi_{\alpha(\delta)} - f^\delta \rangle \\ &= \langle \mathcal{T}(\varphi_{\alpha(\delta)} - \varphi^\dagger), \mathcal{T}\varphi_{\alpha(\delta)} - f^\delta \rangle + \langle f^\dagger - f^\delta, \mathcal{T}\varphi_{\alpha(\delta)} - f^\delta \rangle \\ &= \langle \varphi_{\alpha(\delta)} - \varphi^\dagger, \mathcal{T}^*(\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta) \rangle + \langle f^\dagger - f^\delta, \mathcal{T}\varphi_{\alpha(\delta)} - f^\delta \rangle \\ &\leq \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \|\mathcal{T}^*\| \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| + \delta \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|. \end{aligned}$$

\square

3. Monotonicity of the gradient of convex functionals

If the positive real valued convex functional $\mathcal{P}(\cdot) : \mathcal{C}^1(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$, is in the class of \mathcal{C}^1 , then for all φ, Ψ defined on $\Omega \subset \mathcal{H}$,

$$\mathcal{P}(\Psi) \geq \mathcal{P}(\varphi) + \langle \nabla \mathcal{P}(\varphi), \Psi - \varphi \rangle. \quad (3.1)$$

What this inequality basically means is that at each φ the tangent line of the functional lies below the functional itself. The same is also true from subdifferentiability point of view. Following from (3.1), one can also write that

$$\mathcal{P}(\varphi) - \mathcal{P}(\Psi) \leq \langle \nabla \mathcal{P}(\varphi), \varphi - \Psi \rangle. \quad (3.2)$$

Still from (3.1), by replacing φ with Ψ one obtains

$$\mathcal{P}(\varphi) \geq \mathcal{P}(\Psi) + \langle \nabla \mathcal{P}(\Psi), \varphi - \Psi \rangle, \quad (3.3)$$

or equivalently

$$\mathcal{P}(\varphi) - \mathcal{P}(\Psi) \geq \langle \nabla \mathcal{P}(\Psi), \varphi - \Psi \rangle. \quad (3.4)$$

Combining (3.2) and (3.4) brings us,

$$\langle \nabla \mathcal{P}(\Psi), \varphi - \Psi \rangle \leq \mathcal{P}(\varphi) - \mathcal{P}(\Psi) \leq \langle \nabla \mathcal{P}(\varphi), \varphi - \Psi \rangle. \quad (3.5)$$

Eventually this implies

$$0 \leq \langle \nabla \mathcal{P}(\varphi) - \nabla \mathcal{P}(\Psi), \varphi - \Psi \rangle \quad (3.6)$$

which is the monotonicity of the gradient of convex functionals, [3, Proposition 17.10].

Initially, owing to the relation in (3.5), it can easily be shown the weak convergence of the regularized solution $\varphi_{\alpha(\delta)}$ to the true solution φ^\dagger , with the choice of regularization parameter $\alpha(\delta)$.

THEOREM 3.1 (Weak convergence of the regularized solution). *In the same conditions of Lemma 2.5, if the regularized minimum $\varphi_{\alpha(\delta)}$ to the problem (2.2) exists and $\|f^\delta - f^\dagger\|_{\mathcal{L}^2} \leq \delta$, then*

$$\varphi_{\alpha(\delta)} \rightharpoonup \varphi^\dagger, \text{ as } \alpha(\delta) = \delta^p \rightarrow 0 \text{ for any } p \in (0, 2). \quad (3.7)$$

Proof. Since $\varphi_{\alpha(\delta)}$ is the minimizer of the cost functional $F(\varphi, f^\delta) : \mathcal{H} \rightarrow \mathbb{R}_+$, then

$$F(\varphi_{\alpha(\delta)}, f^\delta) = \frac{1}{2} \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|_{\mathcal{L}^2}^2 + \alpha J(\varphi_{\alpha(\delta)}) \leq \frac{1}{2} \|\mathcal{T}\varphi^\dagger - f^\delta\|_{\mathcal{L}^2}^2 + \alpha J(\varphi^\dagger) = F(\varphi^\dagger, f^\delta),$$

which is in other words,

$$\alpha(J(\varphi_{\alpha(\delta)}) - J(\varphi^\dagger)) \leq \frac{1}{2} \|\mathcal{T}\varphi^\dagger - f^\delta\|_{\mathcal{L}^2}^2 - \frac{1}{2} \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|_{\mathcal{L}^2}^2. \quad (3.8)$$

From the convexity of the penalization term $J(\cdot)$, a lower boundary has been already found in (3.5). Then following from (3.5), the last inequality implies,

$$\alpha \langle \nabla J(\varphi^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle \leq \alpha(J(\varphi_{\alpha(\delta)}) - J(\varphi^\dagger)) \leq \frac{1}{2} \delta^2, \quad (3.9)$$

since $\|f^\dagger - f^\delta\|_{\mathcal{L}^2} \leq \delta$. With the choice of $\alpha(\delta) = \delta^p$ for any $p \in (0, 2)$, desired result is obtained

$$\langle \nabla J(\varphi^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle \leq \frac{1}{2} \delta^{p-2}. \quad (3.10)$$

□

REMARK 3.2. *Note that the result of the theorem is true for any smooth and convex penalty $J(\cdot)$ in the problem (2.2).*

4. Convergence Results for $\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|$

We now come to the point where we analyse each cases when $J(\cdot) : \mathcal{C}^k(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$ for $k \in \{1, 2\}$. In each case, we will consider the discrepancy principle for the choice of regularization parameter while providing the norm convergence.

4.1. When the penalty $J(\cdot)$ is defined over $\mathcal{C}^1(\Omega, \mathcal{H})$

First part of the following formulation has been studied in [6, Theorem 5.]. There, the authors obtain some convergence in terms of a Lagrange multiplier $\lambda(\delta)$ instead of a regularization parameter $\alpha(\delta)$. According to theoretical set up given by the authors, their convergence rate explicitly contain Lagrange multiplier defined as $\lambda(\delta) := 1/\delta$. Second part, on the other hand, has been motivated by [16, Theorem 4.4]. All convergence results are obtained under the assumption that the penalizer is 2-convex according to (2.5).

THEOREM 4.1 (Upper bound for the Bregman divergence associated with the penalty). *Let $J(\cdot) : \mathcal{C}^1(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$ be the smooth and 2-convex penalization term of the cost functional $F(\cdot, f)$ given in the problem (2.2), and denote by $\varphi_{\alpha(\delta)}$ the regularized solution of the same problem satisfying $\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \tau\delta$ where $\tau \geq 1$ as in (2.2). Then, the choice of regularization parameter $\alpha(\delta) := \sqrt{\delta}(\tau + 1)\|\mathcal{T}^*\|$ yields,*

$$D_J(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq \sqrt{\delta} \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|, \quad (4.1)$$

and

$$D_J^{sym}(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq \sqrt{\delta} \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|, \quad (4.2)$$

both of which imply,

$$\|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \leq \sqrt{\delta}. \quad (4.3)$$

Proof. First recall the formulation for the Bregman divergence associated with the penalty $J(\cdot)$ in (2.7). Convexity of the penalizer $J(\cdot)$ brings the following estimation by the second part of (3.5),

$$J(\varphi_{\alpha(\delta)}) - J(\varphi^\dagger) \leq \langle \nabla J(\varphi_{\alpha(\delta)}), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle$$

Then in fact (2.7) can be bounded by,

$$\begin{aligned} D_J(\varphi_{\alpha(\delta)}, \varphi^\dagger) &\leq \langle \nabla J(\varphi_{\alpha(\delta)}) - \nabla J(\varphi^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle \\ &= \frac{1}{\alpha(\delta)} \langle \mathcal{T}^*(f^\delta - \mathcal{T}\varphi_{\alpha(\delta)}) - \mathcal{T}^*(f^\delta - \mathcal{T}\varphi^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle, \end{aligned}$$

due to the first order optimality conditions in (2.4), *i.e.* $\mathcal{T}^*(f^\delta - \mathcal{T}(\cdot)) = \alpha(\delta)\nabla J(\cdot)$. The inner product can also be written in the composite form,

$$D_J(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq \frac{1}{\alpha(\delta)} \langle \mathcal{T}^*(f^\delta - \mathcal{T}\varphi_{\alpha(\delta)}), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle - \frac{1}{\alpha(\delta)} \langle \mathcal{T}^*(f^\delta - f^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle,$$

where the true solution φ^\dagger satisfies $\mathcal{T}\varphi^\dagger = f^\dagger$. Taking absolute value of the right hand side with Cauch-Schwarz inequality and recalling that $\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \tau\delta$ by (2.18) brings

$$\begin{aligned} D_J(\varphi_{\alpha(\delta)}, \varphi^\dagger) &\leq \frac{\tau\delta}{\alpha(\delta)} \|\mathcal{T}^*\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| + \frac{\delta}{\alpha(\delta)} \|\mathcal{T}^*\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \\ &= \frac{(\tau+1)\delta}{\alpha(\delta)} \|\mathcal{T}^*\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|. \end{aligned} \tag{4.4}$$

As for the upper bound for $D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger)$, we adapt (2.15) in the following way

$$\begin{aligned} D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger) &= D_J(\varphi_{\alpha(\delta)}, \varphi^\dagger) + D_J(\varphi^\dagger, \varphi_{\alpha(\delta)}) \\ &= \langle \nabla J(\varphi^\dagger) - \nabla J(\varphi_{\alpha(\delta)}), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle. \end{aligned}$$

Again by the first order optimality conditions in (2.4), then

$$D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger) = \frac{1}{\alpha(\delta)} \langle \mathcal{T}^*(f^\delta - f^\dagger) - \mathcal{T}^*(f^\delta - \mathcal{T}\varphi_{\alpha(\delta)}), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle$$

We split this inner product over the term $\varphi_{\alpha(\delta)} - \varphi^\dagger$ together with the absolute value of each part as such,

$$\begin{aligned} D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger) &\leq \frac{1}{\alpha(\delta)} \left\{ |\langle \mathcal{T}^*(f^\delta - f^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle| + \frac{1}{\alpha(\delta)} |\langle \mathcal{T}^*(f^\delta - \mathcal{T}\varphi_{\alpha(\delta)}), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle| \right\} \\ &\leq \frac{1}{\alpha(\delta)} \left\{ \delta \|\mathcal{T}^*\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| + \|\mathcal{T}^*\| \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \right\}, \end{aligned}$$

which is the consequence of Cauchy-Schwarz. Now again by the condition in (2.18)

$$D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq \frac{1}{\alpha(\delta)} \left\{ \delta \|\mathcal{T}^*\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| + \tau\delta \|\mathcal{T}^*\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \right\}. \tag{4.5}$$

Considering the defined regularization parameter, $\alpha(\delta) := \sqrt{\delta}(\tau + 1)\|\mathcal{T}^*\|$, both in (4.4) and in (4.5) yields the desired upper bounds for $D_J(\varphi_{\alpha(\delta)}, \varphi^\dagger)$ and $D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger)$ respectively. Since $J(\cdot)$ is 2-convex, then the norm convergence of $\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|$ is obtained due to (2.5). \square

In fact those rates also imply another faster convergence rate when the regularization parameter is defined as $\alpha(\delta) := \sqrt{\delta}(\tau + 1)\|\mathcal{T}^*\|$. To observe this, different formulation of the Bregman divergence is necessary. In the Definition 2.1, take $\Phi(\cdot) := G_\delta(\cdot, f^\delta) = \frac{1}{2}\|\mathcal{T}(\cdot) - f^\delta\|^2$ to formulate the following. However, we need to recall the assumptions about the 2-convexity of $G_\delta(\cdot, f^\delta)$ in (2.12) and (2.13).

THEOREM 4.2. *Let $\mathcal{T} : \mathcal{H} \rightarrow \mathcal{H}$ be a compact forward operator in the problem (2.2) and assume that the conditions in (2.12) and (2.13) are satisfied. We formulate a Bregman divergence associated with the misfit term $G_\delta(\cdot, f^\delta) := \frac{1}{2}\|\mathcal{T}(\cdot) - f^\delta\|^2$,*

$$D_G(\varphi_{\alpha(\delta)}, \varphi^\dagger) = \frac{1}{2}\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|^2 - \frac{1}{2}\|\mathcal{T}\varphi^\dagger - f^\delta\|^2 - \langle \nabla G_\delta(\varphi^\dagger, f^\delta), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle. \quad (4.6)$$

If $\varphi_{\alpha(\delta)}$ is the regularized minima for the problem (2.2), then with the choice of regularization parameter $\alpha(\delta) := \sqrt{\delta}(\tau + 1)\|\mathcal{T}^*\|$ for sufficiently small $\delta \in (0, 1)$,

$$D_{G_\delta}(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq \mathcal{O}(\delta^{3/2}) \quad (4.7)$$

As expected, this rate also implies the following

$$\|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \leq \mathcal{O}(\delta^{3/4}). \quad (4.8)$$

Proof. As given by (2.18), $\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \tau\delta$. Additionally the noisy measurement f^δ to the true measurement f^\dagger satisfies $\|f^\delta - f^\dagger\| \leq \delta$. In the Theorem 4.1 above, we have estimated a pair of convergence rates with the same regularization parameter $\alpha(\delta)$. So for $D_{G_\delta}(\varphi_{\alpha(\delta)}, \varphi^\dagger)$ defined by (4.6) will provide the result below;

$$\begin{aligned} D_{G_\delta}(\varphi_{\alpha(\delta)}, \varphi^\dagger) &\leq \frac{1}{2}(\tau\delta)^2 + \frac{1}{2}\delta^2 - \langle \mathcal{T}^*(f^\dagger - f^\delta), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle \\ &= \frac{1}{2}(\tau\delta)^2 + \frac{1}{2}\delta^2 - \langle f^\dagger - f^\delta, \mathcal{T}^*(\varphi_{\alpha(\delta)} - \varphi^\dagger) \rangle \\ &\leq \frac{1}{2}\delta^2(\tau^2 + 1) + \delta\|\mathcal{T}^*\|\|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \end{aligned}$$

As has been estimated in the Theorem 4.1 $\|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \leq \sqrt{\delta}$ when $\alpha(\delta) := \sqrt{\delta}(\tau + 1)\|\mathcal{T}^*\|$. Hence,

$$\begin{aligned} D_{G_\delta}(\varphi_{\alpha(\delta)}, \varphi^\dagger) &\leq \frac{1}{2}\delta^2(\tau^2 + 1) + \delta^{3/2}\|\mathcal{T}^*\| \\ &\leq \delta^{3/2} \left(\frac{1}{2}(\tau^2 + 1) + \|\mathcal{T}^*\| \right). \end{aligned} \quad (4.9)$$

Now, since $G_\delta(\cdot, f^\delta)$ is 2-convex (see Def. 2.1), by (2.5) and by the assumptions (2.12) and (2.13), we have,

$$\|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \leq \frac{\delta^{3/4}}{c_{f^\delta}} \left(\frac{1}{2}(\tau^2 + 1) + \|\mathcal{T}^*\| \right)^{1/2}. \quad (4.10)$$

□

4.2. When the penalty $J(\cdot)$ is defined over $\mathcal{C}^{2+}(\Omega, \mathcal{H})$

Surely the convergence rates above are still preserved when the penalty $J(\cdot)$ is defined over $\mathcal{C}^{2+}(\Omega, \mathcal{H})$ since $\mathcal{C}^2 \subset \mathcal{C}^1$. However, one may be interested in discrepancy principle in this more specific case. Above, we have formulated those convergence rates under the assumption $J(\cdot) : \mathcal{C}^1(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$. We will now analyse the convergence with assuming $J(\cdot) : \mathcal{C}^{2+}(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$. Here we will define regularization parameter also as a function of Hessian Lipschitz constant L_H , [14]. We begin with estimating the discrepancy $\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|_2$.

THEOREM 4.3. *Let $F_\alpha(\cdot)$ be the smooth and convex cost functional as defined in the problem (2.2). If the penalty $J(\cdot) : \mathcal{C}^{2+}(\Omega, \mathcal{H}) \rightarrow \mathbb{R}_+$ is strongly convex, then*

$$\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \delta \left(1 + \frac{1}{L_H} \|\mathcal{T}^*\|^2 \right)^{1/2} + \sqrt{\tilde{\mathcal{O}}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2)}$$

where L_H is the Hessian Lipschitz constant of the functional $F_\alpha(\cdot)$.

Proof. Let us consider the following second order Taylor expansion,

$$\begin{aligned} F_\alpha(\varphi_{\alpha(\delta)}) &= F_\alpha(\varphi^\dagger) + \langle \varphi_{\alpha(\delta)} - \varphi^\dagger, \nabla F_\alpha(\varphi^\dagger) \rangle + \\ &\quad + \frac{1}{2} \langle \varphi_{\alpha(\delta)} - \varphi^\dagger, \nabla^2 F_\alpha(\varphi^\dagger) (\varphi_{\alpha(\delta)} - \varphi^\dagger) \rangle + \mathcal{O}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2) \end{aligned}$$

Obviously, this Taylor expansion is bounded by

$$F_\alpha(\varphi_{\alpha(\delta)}) \leq F_\alpha(\varphi^\dagger) + \langle \varphi_{\alpha(\delta)} - \varphi^\dagger, \nabla F_\alpha(\varphi^\dagger) \rangle + \frac{1}{2} L_H \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2 + \mathcal{O}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2),$$

where L_H is the Hessian Lipschitz constant of the functional $F_\alpha(\cdot)$. After some arrangement with the explicit definition $F_\alpha(\cdot)$ in the problem (2.2) the inequality above reads,

$$\begin{aligned} \frac{1}{2} \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|^2 &\leq \frac{1}{2} \delta^2 + \alpha(\delta) (J(\varphi^\dagger) - J(\varphi_{\alpha(\delta)})) + \\ &\quad + \alpha(\delta) \langle \varphi_{\alpha(\delta)} - \varphi^\dagger, \nabla J(\varphi^\dagger) \rangle + \langle \varphi_{\alpha(\delta)} - \varphi^\dagger, \mathcal{T}^*(f^\dagger - f^\delta) \rangle + \frac{1}{2} L_H \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2 \\ &\quad + \mathcal{O}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2). \end{aligned}$$

Now by the early estimations for the difference $J(\varphi^\dagger) - J(\varphi_{\alpha(\delta)})$ in (3.2),

$$\begin{aligned} \frac{1}{2} \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|^2 &\leq \frac{1}{2}\delta^2 + \alpha(\delta) \langle \nabla J(\varphi^\dagger), \varphi^\dagger - \varphi_{\alpha(\delta)} \rangle - \alpha(\delta) \langle \nabla J(\varphi^\dagger), \varphi^\dagger - \varphi_{\alpha(\delta)} \rangle + \\ &+ \langle \varphi_{\alpha(\delta)} - \varphi^\dagger, \mathcal{T}^*(f^\dagger - f^\delta) \rangle + \frac{1}{2}L_H \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2 + \mathcal{O}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2). \end{aligned}$$

After Cauchy-Schwarz and Young's inequalities on the right hand side, we have

$$\begin{aligned} \frac{1}{2} \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|^2 &\leq \frac{1}{2}\delta^2 + \delta \|\mathcal{T}\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| + \frac{1}{2}L_H \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2. \\ &\leq^\dagger \delta^2 \left(\frac{1}{2} + \frac{1}{2L_H} \|\mathcal{T}\|^2 \right) + L_H \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2 + \mathcal{O}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2). \end{aligned}$$

In the name of convenience, we combine the last two terms on the right hand side under one notation $\tilde{\mathcal{O}}$. Then,

$$\frac{1}{2} \|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\|^2 \leq \delta^2 \left(\frac{1}{2} + \frac{1}{2L_H} \|\mathcal{T}\|^2 \right) + \tilde{\mathcal{O}}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2).$$

Since $(\sqrt{a} + \sqrt{b})^2 = a + b + 2\sqrt{ab} \geq a + b$ for $a, b \in \mathbb{R}_+$, hence

$$\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \delta \left(1 + \frac{1}{L_H} \|\mathcal{T}^*\|^2 \right)^{1/2} + \sqrt{\tilde{\mathcal{O}}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2)}. \quad (4.11)$$

□

REMARK 4.4 (The coefficient τ in the limit sense). *In the theorem, the remaining term is $\sqrt{\tilde{\mathcal{O}}(\|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2)}$. As a result of any regularization strategy, it is expected that $\|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \rightarrow 0$ as $\alpha(\delta) \rightarrow 0$. Hence in the limit sense, the coefficient τ in (2.18) may be defined as*

$$\tau(L_H) := \left(1 + \frac{1}{L_H} \|\mathcal{T}^*\|^2 \right)^{1/2}. \quad (4.12)$$

REMARK 4.5 (Preservation of the convergence rates). *Owing to the Theorem 4.1, it is easy to conclude that the convergence rates defined above are preserved when the penalty $J(\cdot)$ is 2-convexity by (2.5) and the regularization parameter is defined as,*

$$\alpha(\delta) := \sqrt{\delta}(\tau(L_H) + 1) \|\mathcal{T}^*\|, \quad (4.13)$$

where $\tau(L_H) := \left(1 + \frac{1}{L_H} \|\mathcal{T}^*\|^2 \right)^{1/2}$.

† For some $\epsilon > 0$, by Young's inequality $\delta \|\mathcal{T}^*\| \|\varphi_{\alpha(\delta)} - \varphi^\dagger\| \leq \frac{\epsilon}{2} \delta^2 \|\mathcal{T}^*\|^2 + \frac{1}{2\epsilon} \|\varphi_{\alpha(\delta)} - \varphi^\dagger\|^2$. If we take $\epsilon = 1/L_H$, then the inequality follows.

5. Summary of the Convergence Rates

In this work, we have obtained the convergence rates with following the footsteps of the counterpart works in [6, 15, 16]. However, we have also taken into account one more fact which is $\|\mathcal{T}\varphi_{\alpha(\delta)} - f^\delta\| \leq \tau\delta$ where $\alpha(\delta)$ fulfils the condition (2.18). It has been observed that 2-convexity condition for the penalty $J(\cdot)$ is crucial to obtain norm convergence by means of Bregman divergence. We have not given any analytical evaluation of τ without any specific penalty $J(\cdot)$. Note that these convergence rates are true for $J(\cdot) \in \mathcal{C}^k(\Omega, \mathcal{H})$ where $k = 1$ and $k = 2$. Below we summarize these corresponding convergence rate estimations per Bregman divergence formulation.

$\alpha(\delta)$	Bregman divergence estimate	$\ \varphi_{\alpha(\delta)} - \varphi^\dagger\ $ estimate
$\alpha(\delta) := \sqrt{\delta}(\tau + 1)\ \mathcal{T}^*\ $	$D_J(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq \mathcal{O}(\sqrt{\delta}),$	$\sqrt{\delta}$
$\alpha(\delta) := \sqrt{\delta}(\tau + 1)\ \mathcal{T}^*\ $	$D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq \mathcal{O}(\sqrt{\delta}),$	$\sqrt{\delta}$
$\alpha(\delta) := \sqrt{\delta}(\tau + 1)\ \mathcal{T}^*\ $	$D_G(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq \mathcal{O}(\delta^{3/2}),$	$\frac{\delta^{3/4}}{c_f\delta} \left(\frac{1}{2}(\tau^2 + 1) + \ \mathcal{T}^*\ \right)^{1/2}$

Acknowledgement

The author is indepted to Prof. Dr. D. Russell Luke for valuable discussions on different parts of this work.

A. Further properties of the Bregman divergence

Although $D_F(\varphi_{\alpha(\delta)}, \varphi^\dagger)$ has been introduced above in Definition 2.2 by (2.6), an immediate conclusion can be formulated below.

COROLLARY A.1. *If $\varphi_{\alpha(\delta)}$ and φ^\dagger are the regularized and the true solutions respectively to the problem (2.2) wherein the cost functional F is convex and smooth, then*

$$D_F(\varphi_{\alpha(\delta)}, \varphi^\dagger) = 0. \quad (1.1)$$

Proof. Since $\varphi_{\alpha(\delta)}$ is the minimizer, then $F(\varphi_{\alpha(\delta)}) \leq F(\varphi)$ for all $\varphi \in \Omega$, which implies

$$D_F(\varphi_{\alpha(\delta)}, \varphi^\dagger) \leq -\langle \nabla F(\varphi^\dagger), \varphi_{\alpha(\delta)} - \varphi^\dagger \rangle. \quad (1.2)$$

On the other hand, just by logic, $\nabla F(\varphi^\dagger) = 0$. It is known that, for any convex functional Φ the Bregman divergence $D_\Phi \geq 0$. Hence $D_F(\varphi_{\alpha(\delta)}, \varphi^\dagger) = 0$. □

Addition to this, a relation between $D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger)$ and $D_G^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger)$ can also be observed.

THEOREM A.2. *Let the regularized minimum $\varphi_{\alpha(\delta)}$ to the problem (2.2) satisfy the first order optimality conditions (2.4). Then the following inclusion holds true for $\alpha > 0$,*

$$\alpha D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger) = D_G^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger). \quad (1.3)$$

Proof. As defined by (2.8), one can directly derive

$$\begin{aligned} D_G^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger) &= \langle \nabla G(\varphi_\alpha^\delta) - \nabla G(\varphi^\dagger), \varphi_\alpha^\delta - \varphi^\dagger \rangle \\ &= \langle \mathcal{T}^*(\mathcal{T}\varphi_\alpha^\delta - f^\delta) - \mathcal{T}^*(f^\dagger - f^\delta), \varphi_\alpha^\delta - \varphi^\dagger \rangle \end{aligned} \quad (1.4)$$

In proof of Theorem 4.1, or by (2.15), $D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger)$ has been given already. Since φ_α^δ satisfies the first order optimality conditions (2.4),

$$D_J^{\text{sym}}(\varphi_{\alpha(\delta)}, \varphi^\dagger) = \frac{1}{\alpha} \langle \mathcal{T}^*(\mathcal{T}\varphi_\alpha^\delta - f^\delta) - \mathcal{T}^*(f^\dagger - f^\delta), \varphi_\alpha^\delta - \varphi^\dagger \rangle \quad (1.5)$$

which yields the result. □

References

- [1] R. Acar, C. R. Vogel. *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, Vol. 10, No. 6, 1217 - 1229, 1994.
- [2] M. Bachmayr and M. Burger. *Iterative total variation schemes for nonlinear inverse problems*, Inverse Problems, 25, 105004 (26pp), 2009.
- [3] H. H. Bauschke, P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, Springer New York, 2011.
- [4] J. M. Bardsley and A. Luttmann. *Total variation-penalized Poisson likelihood estimation for ill-posed problems*, Adv. Comput. Math., 31:25-59, 2009.
F. Schöpfer, T. Schuster. *Minimization of Tikhonov Functionals in Banach Spaces*, Abstr. Appl. Anal., Art. ID 192679, 19 pp, 2008 .
- [5] K. Bredies. *A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space*, Inverse Problems 25, no. 1, 015005, 20 pp, 2009.
- [6] M. Burger, S. Osher. *Convergence rates of convex variational regularization*, Inverse Problems, 20(5), 1411 - 1421, 2004.
- [7] A. Chambolle, P. L. Lions. *Image recovery via total variation minimization and related problems*, Numer. Math. 76, 167 - 188, 1997.
- [8] T. F. Chan and K. Chen. *An optimization-based multilevel algorithm for total variation image denoising*, Multiscale Model. Simul. 5, no. 2, 615-645, 2006.
- [9] T. Chan, G. Golub and P. Mulet. *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comp 20: 1964-1977, 1999.
- [10] D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer Verlag Series in Applied Mathematics Vol. 93, Third Edition 2013.
- [11] D. Dobson, O. Scherzer. *Analysis of regularized total variation penalty methods for denoising*, Inverse Problems, Vol. 12, No. 5, 601 - 617, 1996.
- [12] D. C. Dobson, C. R. Vogel. *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal., Vol. 34, No. 5, 1779 - 1791, 1997.
- [13] H. W. Engl, M. Hanke, A. Neubauer. *Regularization of inverse problems*, Math. Appl., 375. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [14] J. M. Fowkes, N. I. M. Gould, C. L. Farmer. *A branch and bound algorithm for the global optimization of Hessian Lipschitz continuous functions*, J. Glob. Optim., 56, 1792 - 1815, 2013.
- [15] M. Grasmair. *Generalized Bregman distances and convergence rates for non-convex regularization methods*, Inverse Problems 26, 11, 115014, 16pp, 2010.
- [16] M. Grasmair. *Variational inequalities and higher order convergence rates for Tikhonov regularisation on Banach spaces*, J. Inverse Ill-Posed Probl., 21, 379-394, 2013.
- [17] M. Grasmair, M. Haltmeier, O. Scherzer. *Necessary and sufficient conditions for linear convergence of l^1 -regularization*, Comm. Pure Appl. Math. 64(2), 161-182, 2011.
- [18] V. Isakov. *Inverse problems for partial differential equations*. Second edition. Applied Mathematical Sciences, 127. Springer, New York, 2006.
- [19] A. Kirsch. *An introduction to the mathematical theory of inverse problems*. Second edition. Applied Mathematical Sciences, 120. Springer, New York, 2011.
- [20] D. A. Lorenz. *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, J. Inv. Ill-Posed Problems, 16, 463-478, 2008.
- [21] A. N. Tikhonov. *On the solution of ill-posed problems and the method of regularization*, Dokl. Akad. Nauk SSSR, 151, 501-504, 1963.
- [22] A. N. Tikhonov, V. Y. Arsenin. *Solutions of ill-posed problems*. Translated from the Russian. Preface by translation editor Fritz John. Scripta Series in Mathematics. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York-Toronto, Ont.-London, xiii+258 pp, 1977.
- [23] R.T. Rockafellar, R. J.-B. Wets. *Variational Analysis*. Fundamental Principles of Mathematical Sciences, 317. Springer-Verlag, Berlin, 1998.

- [24] L. I. Rudin, S. J. Osher, E. Fatemi. *Nonlinear total variation based noise removal algorithms*, Physica D, 60, 259-268, 1992.
- [25] C. R. Vogel, M. E. Oman. *Iterative methods for total variation denoising*, SIAM J. SCI. COMPUT., Vol. 17, No. 1, 227-238, 1996.